DOCUMENT RESUME

ED 275 174 FL 016 078

AUTHOR Sussmilch, Edgar
TITLE Language Testing with Immigrant Children.
PUB DATE 84
NOTE 12p.; In: Practice and Problems in Language Testing. Papers from the International Symposium on Language Testing (7th, Colchester, England, 1984); see FL 016 066.
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Children; Cloze Procedure; Foreign Countries; *German; *Immigrants; Language Proficiency; *Language Tests; *Second Languages; *Standardized Tests; Student Placement; *Test Use
IDENTIFIERS *C Tests; West Germany

ABSTRACT

This study examines the differences of approaches to second language testing in West Germany in regard to linguistic bases, pedagogical aims, and the desired degree of standardization and empirical quality control. One approach demands a curriculum-integrated diagnosis of language development aiming at a highly differentiated description of the individual second language proficiency level to determine appropriate therapeutic procedures. A second approach demands the development of standardized language tests that are objective, reliable, and valid. Investigation of the C-Test suggests that it fills all of the requirements of the second approach, and is useful in classroom testing, selection, classification, and placement. The lack of suitable standardized tests for second language learners in Germany has stimulated development of informal testing procedures, but most have not been properly researched and are cumbersome to administer. Standardized tests such as the C-Test have the advantage of economy of administration and interpretation, and could play an important role in improving the educational opportunities of immigrant children in Germany. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED275174

LANGUAGE TESTING WITH IMMIGRANT CHILDREN

EDGAR SÜßMILCH
(UNIVERSITY OF DÜSSELDORF)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

P. Culhane

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

FL 016 078

LANGUAGE TESTING WITH IMMIGRANT CHILDREN

Edgar Sußmilch (University of Dusseldorf)

1 The educational situation of immigrant children in West Germany

The number of non-German pupils in the Federal Republic of Germany and West-Berlin rose steeply in the Seventies. Recruitment of foreign workers ceased in 1973 but many children have since entered the country to join their parents. In addition many children of immigrant workers have been born in Germany. In the school year 1981-82 816 000 non-German children were attending schools in the Federal Republic. In the Grundschulen (primary schools) and the Hauptschulen (secondary modern schools) the proportion of non-Germans reached 12.6% overall, in North-Rhine-Westphalia 15.3%, and in Berlin, which has the highest proportion of immigrant pupils, every fourth child in the non-selective schools was a foreigner. However, because of the unequal distribution over the country in some cities and city areas the proportions are even higher.

Only recently and very gradually have the individual federal states begun to develop concepts and strategies to improve the educational provisions for these children and adolescents, and one of the undecided issues is the relative importance of the native language and of German as a second language. It is obvious that German is a necessary precondition for success in school and employment opportunities, as well as for social and cultural integration in general. But teaching German under these conditions is a very difficult enterprise. First of all, the teachers have not been trained to perform this task. There are few suitable text books available. Classes are large, and the non-German children vary greatly in their command of the German language. However, since remedial language teaching must take this heterogenity as its starting point, an objective technique for determining proficiency levels would be useful.

2 Measuring proficiency in German as a second language

2.1 Language diagnosis and language training

Tests in the school setting are instruments for acquiring information to be used in planning teaching strategies and evaluating teaching success. They have an important contribution to make as objective assessment instruments to help the teacher in making judgements about his pupils. For the immigrant children tests could support the teacher in making classification decisions on school entry as well as in determining whether pupils should be assigned to preparatory or to normal classes. A further decision point where language tests could be useful is at the end of the fourth grade where pupils can be sent either to a selective secondary school or to the Hauptschule. Language tests can also be used for diagnosis and treatment decisions.

In the last few years a number of test development projects have been undertaken. They have a variety of different linguistic foundations and pedagogical objectives. Because there are at present virtually no suitable tests commercially available, many approaches have been tried by practising teachers in order to provide a better basis for evaluating the German proficiency of the immigrant children and for developing appropriate teaching strategies. Often the aim has been to combine diagnosis and therapy. However, in many cases these tests do not proceed beyond initial trials, and empirical investigation and analysis are not performed so that such procedures remain questionable both linguistically and psychometrically. A brief survey of the available approaches will show this more clearly.

2.2 Informal language tests

Ihssen, speaking about the methodological problems of language diagnosis with immigrant children, demands "an account of the individual language biography and the sociolinguistic conditions ... in order to obtain an adequate assessment of the developmental level in the first and second language, for parental counselling, for the planning, implementation and evaluation of supporting measures" (1980:40). As the primary source of information for a diagnosis of language development he proposes the analysis of spontaneous language behaviour in a natural communication situation. To do this he takes a sample (representative if possible) of 50 consecutive sentences. Using a grid taken from child language development research he analyses word and sentence structure (morphology and syntax). His categories include, for instance, the average number of morphemes per sentence, wrongly used grammatical morphemes and word order errors.

Many other approaches follow Ihssen's lead. Informal tests are developed which cover only a limited segment of language derived from a structural analysis. Obviously it is felt that such approaches are particularly suited for therapeutic purposes. Standardised (group) tests which comply with the normal test requirements of objectivity, reliability and validity are rejected, since it is felt that they are incompatible with the therapeutic approach. Instead tests are developed which cannot be evaluated by the traditional techniques and which, therefore, are of dubious value.

Scherzinger and Scherzinger (1981) expand Ihssen's approach by using picture stories as elicitation techniques. This permits the collection of comparable data for each subject. The same picture sequence is used to elicit an oral and a written sample. Here again only morphosyntactical features are scored, although qualitative criteria, such as breadth of vocabulary and communicative competence, are mentioned.

Other test projects also stress the analysis of morphosyntactical features in spontaneous speech samples: obviously it is felt that a structural model of language is particularly suitable for the assessment of L2 proficiency with immigrant children (cf. Portz and Pfaff 1981; Steinmuller 1983). A technique for analysing written texts is offered by Hegele (1981). The aim of her investigation is to provide baseline data for planning therapeutic measures and for

evaluating their effectivity.

In three further projects the attempt has been made to provide extensive item pools and to provide information about test administration and scoring. However, even these are informal tests since they do not conform to the accepted test standards, which reduces their usefulness. In addition item analyses have not been attempted so that improvement of the individual items is not possible.

In the first project the test for transfer into the normal 4th grade class was produced by the Akademie für Lehrerfortbildung Dillingen (no date). It was the first extensive language test developed for immigrant children. It consists of an individual interview with tests of free speech production, guided speech production, listening comprehension, reading comprehension and a written group test with the subtests vocabulary, grammatical structure and dictation. No details are provided about the total time needed for test administration. A first trial with 34 subjects showed a wide range of scores.

Luchtenberg, Neumann, and Wespel (1982) also developed an informal test. It is restricted to oral communicative ability: understanding simple instructions, simple questions and statements, active vocabulary, simple word and sentence structures in German. For this individual test (which can take up to two hours to administer) the only data reported consist of a fictitious results table for 5 subjects.

Fliegner and Gogolin (1982) provide a test to be used with school beginners. Revised material for the areas listening comprehension and speaking, split into word, sentence and text level, is accompanied by a teachers' manual. The materials provided form the basis for a 20 minute individual interview. However, no controlled trials have been conducted and the revision was performed on the basis of critical suggestions and teacher reports. An empirical investigation of the materials is planned and will be conducted by Boos-Nunning in Essen.

2.3 Standardised language tests

In a number of studies language tests developed for German L1 speakers have been administered to immigrant children. Mitzschke (1978) used the reading comprehension test VL 5-6 for a group of non-German youngsters involved in pre-apprenticeship training (N = 37). Damanakis (1978) also used available German tests: the subtests Vocabulary and Reading Comprehension from the AST 2 for Greek pupils (N = 49) in preparatory classes in the second grade. In a second study, concerned with L1 and L2 development in Greek pupils attending normal classes in the 4th and 6th grade (N = 114), Damanakis used the vocabulary test WST 4-6 and the orthography test RST 4+ together with two Greek tests he developed himself. Both he and Mitzschke compared their results with those of the German norming samples, although the comparison is restricted to means and standard deviations.

Biehl (1982) has reported the results of a pilot study performed

in the framework of a Duisburg project investigating attained language proficiency and conditions of language learning. Tape recordings were made of an interview and of a picture story narrative, and newly developed vocabulary and syntax tests were administered. For the vocabulary test 60 words were selected from a 1 000-word basic vocabulary list and L1 translations were combined with L1 distractors to produce multiple choice items. The syntax test consists of pictures accompanied by sentences of varying syntactical complexity. The subject has to state who is doing what in the pictures, and in order to assign the correct names he must understand the accompanying sentence. Yet again, for this study involving 75 subjects no data about test analysis are available, which is rather surprising in view of the fact that Duisburg is known to have its fair share of experts in the field of language testing. Interesting relationships between the language data and the social data have been reported by Biehl, but since not enough information has been presented about the tests themselves, it proves impossible to evaluate the validity of the results obtained.

Sußmilch (1981) developed a vocabulary test for immigrant children in the 2nd and 3rd grades. The test aims at evaluating a subcomponent of a superordinate language competence, namely passive vocabulary, which is very important in learning a second language. Test construction began with a survey of available text books to develop a basic vocabulary list for immigrant children. For item construction nouns, verbs and adjectives with different frequencies were used in gap-filling and matching items accompanied by line drawings. The test combines 12 items for each part of speech and it takes 30 minutes.

This test was examined in empirical trials. In all, 426 non-German and 195 German pupils were tested in 29 classes from the Dusseldorf region. The test turned out to be rather easy for the immigrant children. The mean score was 27.1 (highest possible score 36) with a standard deviation of 7.0. For the immigrant children the test had a reliability of r_{tt} = .90 (alpha) and for the German pupils .82. The test is therefore a highly reliable measure of passive vocabulary, although the reliability sinks for the more advanced German group. The validity was assessed on the basis of teacher ratings of vocabulary knowledge and was calculated as r_{tc} = .41 for the immigrant group and .32 for the Germans. Therefore, the validity is not high and sinks further for the German group as the result of an obvious ceiling effect.

2.4 The C-Test in L2 testing

As I have shown the situation in L2 testing in the Federal Republic of Germany at the present time is most unsatisfactory. On the one hand there is an urgent need for economical, yet accurate tests, on the other hand there is no sign that any of the tests currently under development for use in L2 testing will be available for general use in a form fulfilling the normally accepted psychometric criteria in the near future. So, when I first heard the results Klein-Braley and Raatz were obtaining with the C-Test for their L1 and LF populations (cf. Klein-Braley and Raatz, this volume, and

Raatz and Klein-Braley, 1982), I decided to examine the performance of the C-Test for L2 groups. Raatz and Klein-Braley (1983) had already reported positive results in a pilot study, but there only 13 L2-subjects had been involved, so the results were suggestive but not concensive.

In the course of my investigations C-Tests were administered to a number of different immigrant groups. In my first investigation a total of 476 immigrant pupils, namely Greeks and Turks, and 197 German pupils took the tests. The test form used (D3A) had proved to be reliable and valid for German pupils (see Appendix). The C-Test was administered at the beginning of the school year 1982-83. The time limit was 40 minutes. For the non-German children the test proved rather difficult, and this was shown particularly by failure to finish the test. In the third grade, on average, 25 blanks were not attempted at the end of the test, in the fourth grade 10 on average. The mean difficulty for the immigrant children was P = 21 in the third and P = 40 in the fourth grade. Even so, all five test parts had high discrimination indices and the scores ranged from 0 to 83 for the immigrant children and from 1 to 98 for the Germans (out of 100 possible points). The test therefore differentiates well between individuals; nevertheless, it is obviously far too difficult for the 3rd grade pupils. Great differences emerged in the mean scores between different classes: the range lies between 6.3 and 37.6 in the third grade and from 16.3 to 65.0 in the fourth. Table 1 shows the reliability and validity coefficients (investigations 22 and 23).

Table 1

Reliability and validity coefficients
of a German C-Test with immigrant and German children

investigation	nationality	grade	N	r_{tt}	r_{tc}
	immigrants	3	231	.86	.72
	Turks	3	92	.85	.61
22	Greeks	3	107	.85	.74
	other	3	32	.83	-
	Germans	3	109	.92	.80
	immigrants	4	245	.92	.69
	Turks	4	94	.87	.41
23	Greeks	4	96	.92	.82
	other	4	55	.88	.73
	Germans	4	88	.88	.82

Values are given both for the total immigrant sample and for the individual subsamples (nationalities) per grade. The values of the German group permit comparisons to be made between L1 and L2 learners. As Table 1 shows the C-Test is highly reliable and valid. There are only slight differences between the values for the various

nationalities. The test is slightly less reliable for the immigrant children in the third grade and for the German children in the fourth grade. This reflects the fact that in the first case it was too difficult and in the second slightly too easy.

The validation criteria consisted of rating scales ranging from -5 to +5 for written and spoken German. The teacher was asked to use the entire scale to rate his/her pupils. The results were then averaged for each pupil and the validity coefficient for each class was calculated separately. For this reason the values in the column r_{tc} in Table 1 are the median values. Groups smaller than N = 8 were excluded so that no validity coefficients are given there. The other coefficients show that the C-Test is highly valid for all pupils with the exception of the Turkish pupils in the 4th grade. Here, however, the teachers involved in three classes had only known their pupils for less than eight weeks, and did not really feel competent to give an assessment of their pupils' proficiency in German. An additional factor with the Turkish pupils is the low variance in the test scores.

These investigations show that C-Tests can be used to assess language proficiency in German for foreign pupils and that the test is reliable and valid. The test can, in fact, be shortened to only 3 texts without serious losses in reliability. However, the three texts thus selected would be different for the different samples, and this would mean that test development procedures in the traditional fashion would be necessary.

In three further investigations two newly developed C-Tests were used. Test form DG is similar to DJA, but here the number of missing letters was indicated by providing a dash for each one. This, as Klein-Braley had shown in a pilot study, made the test easier. In investigations 24 and 25 this C-Test was administered to approximately half the Greek pupils involved in investigations 22 and 23. The average difficulty for this test was P = 52 in the third grade and P = 69 in the fourth. Table 2 shows the results of these investigations.

Table 2

Reliability and validity coefficients
of German C-Tests with immigrant children

investigation	nationality	grade	N	r_{tt}	r_{tc}
24	Greeks	3	52	.82	.58 - .74
25	Greeks	4	50	.85	.66 - .90
26	Turks	2	41	.94	.72 - .88

The reliability coefficients of r_{tt} = .82 (investigation 24) and .85 (investigation 25) are satisfactory, and again they can be

improved by discarding the less discriminating texts. The validity criterion was the same teacher judgements that had been obtained in the course of investigations 22 and 23, and Table 2 shows the highest and lowest validity coefficient obtained for the individual classes. Test form DG was an entirely new test using texts which had not been tried out before. It is therefore interesting to compare the results obtained for the two C-Tests administered four months before to the same subjects. The correlations for the two sets of scores are $r = .81$ for the third grade and $r = .90$ for the fourth grade. The two C-Tests are therefore parallel tests although there was a time lapse of four months between the administrations, item analysis had not been performed, the deletion technique had been modified and entirely different texts were involved in the two investigations!

For investigation 26 the attempt was made to reduce text difficulty still further. The texts were selected from readers for the first grade or from L2 text books for the first year of teaching. The texts were shorter and had only ten deletions which affected every third word. Again a dash indicated each missing letter. Test form DT contained 6 texts of this type and was administered to 41 Turkish pupils towards the end of the second grade. They were given 60 minutes to complete the test, the first pupil finished in 13 minutes and, on average, they needed 40 minutes. Table 2 shows the results of this investigation (26).

Text difficulty ranges from $P = 26$ to $P = 55$, the mean difficulty is $P = 40$. Test reliability is $r_{tt} = .94$, the highest reliability coefficient so far obtained for a C-Test. Even when the test is shortened to only 3 texts (30 deletions) the reliability drops slightly to $r'_{tt} = .91$. Again teacher ratings were used for validation, and the values for the three classes range from $r_{tc} = .72$ to .88, which shows very high agreement between teacher judgements and C-Test scores.

The investigations into the performance of C-Test for L2 learners in Germany have produced highly satisfactory results. Three different test versions were used and each was shown to be reliable and valid. Once again it has proved possible to develop highly reliable and valid tests with no preliminary test development procedures. The C-Test results show high agreement with teacher ratings. In view of the extreme heterogenity of the L2 groups in German schools it would seem desirable to produce test versions for specific target groups. In particular, extremely easy tests are needed for the early stages of L2 learning, but as has been shown in investigation 26, modification of the C-principle enables the construction of suitable tests for these subjects.

3 Proficiency testing with L2 groups in Germany

The approaches currently available in the Federal Republic of Germany differ not only in their linguistic basis and their pedagogical aims, but also in the degree of standardisation and empirical quality control desired. One approach demands a curriculum-integrated diagnosis of language development aiming at a highly differentiated

description of the individual level of proficiency in L2 in order to determine appropriate therapeutic procedures. A second approach demands the development of standardised language tests which are objective, reliable and valid. As the investigations reported here show the C-Test fulfils all the requirements of this second approach. The C-Test can be used by teachers in normal classroom test procedures, but it can also be used as an aid in selection, classification and placement decisions.

The lack of suitable standardised tests for L2 groups in Germany has provoked the development of informal test procedures in this area by practitioners working in the field. Unfortunately, the majority of these procedures have not been subsequently examined to determine their test qualities. Furthermore, tests for which no details are available regarding item characteristics, administration and scoring objectivity, reliability and validity must be regarded as dubious procedures. The majority of these informal procedures have the additional disadvantage that they must be administered individually and can take up to two hours.

Standardised language tests such as the C-Test on the other hand have the immense advantage that they are economical. The C-Test can be shortened so that only 20 to 30 minutes are needed for test administration, thereby facilitating its use even in the second grade, yet test reliability remains satisfactory. Placement decisions could be made very quickly using the C-Test, and since one of the main problems in teaching L2 learners is the setting up of homogeneous groups for more effective language instruction, the C-Test could play a very important role in improving the educational chances for immigrant children.

Appendix

Examples of C-Tests

1 C-Test D3A (text 3)

Ein junger Hase, der zum erstenmal in die Sonne kam, sah plötzlich seinen Schatten neben sich. Hals üb_____ Kopf ran_____ der Ha_____ davon, do_____ das schw_____ Tier m_____ den lan_____ Hornern ran_____ ebenso sch_____ neben i_____ her. D_____ Hase sch_____ Haken na_____ rechts u_____ links u_____ jagte i_____ den dun_____ Wald zur_____. Da end_____ war se_____ Verfolger verschwunden. Atemlos und erschopft hielt der Hase an.

2 C-Test DG (text 1)

Das kleine Mädchen wollte so gern einmal mit dem Vater wippen. Es

hat sich auf die eine Seite von der Wi _ _ _ gesetzt, u _ _ der Va _ _ _ hat si _ _ auf d _ _ andere Se _ _ _ von d _ _ Wippe ges _ _ _ _ . Da w _ _ das kle _ _ _ Madchen ga _ _ hoch ob _ _ , und d _ _ Vater w _ _ ganz un _ _ _ . "Los! L _ _ !" hat d _ _ kleine Mad _ _ _ _ gerufen. Ab _ _ die Wi _ _ _ konnte nicht wippen. Das kleine Madchen war doch so leicht, und der Vater war so schwer.

3 C-Test DT (text 6)

Im Garten steht ein großer Kirschbaum. Die Kirschen sind reif. Vi _ _ _ Spatzen sitzen i _ den Zweigen. S _ _ picken die Kirs _ _ _ _ an und mac _ _ _ ein Geschrei. D _ _ Nachbar kommt a _ _ dem Haus. E _ schimpft: Diese fre _ _ _ _ Spatzen! Ich m _ _ die Kirschen heute noch pflucken, sonst sind sie alla angepickt.

Bibliography

Akademie für Lehrerfortbildung Dillingen (Hrsg). No date. Deutschunterricht fur Kinder auslandischer Arbeitnehmer. Ein Fortbildungsmodell. Donauwörth: Auer.

Biehl, J. 1982. Sprachstand und Sprachlernbedingungen türkischer Schuler. Unveröffentlichtes Manuskript. Duisburg: Universitat Duisburg Gesamthochschule.

Damanakis, M. 1978. Sozialisationsprobleme der griechischen Gastarbeiterkinder in den Grund- und Hauptschulen des Bundeslandes Nordrhein-Westfalen. Kastellaun: Henn.

Fliegner, J., and I. Gogolin in Zusammenarbeit mit R. Urbanek. 1982. Sprachstandsmessung bei Schulanfangern. Projektberichte des Padagogischen Instituts der Landeshauptstadt Dusseldorf, Heft 5. Neufassung nach Überarbeitung. Dusseldorf: Padagogisches Institut.

Hegele, I. 1981. "Deutsch als Zweitsprache: Sprachstandsdiagnose und Sprachförderung von Auslanderkindern". Deutsch lernen 6: 41-56 (Heft 4).

Ihssen, W.B. 1980. "Probleme der Sprachentwicklungsdiagnose bei Auslanderkindern. Hinweise auf Voraussetzungen zuverlässiger Ergebnisse". Praxis Deutsch, Sonderheft 1980 - Deutsch als Zweitsprache: 40-42.

Luchtenberg, S., Neumann, H.-J., and M. Wespel. 1982. "Informeller Test zur Sprachstandsdiagnose bei ausländischen Kindern und

Jugendlichen". In Turkische Kinder in unseren Schulen. Eine padagogische Herausforderung. Mit Beitragen von U. Coburn-Staege ... Stuttgart: Klett. 134-176.

Mitzschke, M. 1978. "Probleme der Eingruppierung auslandischer Jugendlicher im Deutschunterricht. Ein Praxisbericht". Diskussion Deutsch 9: 77-81.

Portz, R., and C. Pfaff. 1981. SES - Soziolinguistisches Erhebungsinstrument zur Sprachentwicklung. Berlin: Padagogisches Zentrum Berlin.

Raatz, U., and C. Klein-Braley. 1982. "The C-Test: a modification of the cloze procedure". In Culhane, T., Klein-Braley, C., and D.K. Stevenson (eds). 1982. Practice and problems in language testing 4. University of Essex. Occasional papers 26. Colchester: University of Essex Press. 113-138.

- 1983. "Ein neuer Ansatz zur Messung der Sprachleistung. Der C-Test: Theorie und Praxis". In Horn, R., Ingenkamp, K., and R.S. Jager (Hrsg). 1983. Tests und Trends 3. Jahrbuch der Padagogischen Diagnostik. Weinheim: Beltz. 107-138.

Scherzinger, A., and G. Scherzinger. 1981. "Sprachstand-Diagnose fur die schulische Forderung von Auslanderkindern". Auslanderkinder in Schule und Kindergarten 2: 18-22 (Heft 1).

Steinmuller, U. 1983. "Sprachstandserhebungen in Berlin. Bericht uber zwei Projekte". In Regionale Arbeitsstellen zur Forderung auslandischer Kinder und Jugendlicher (Hrsg). 1983. Sprachstandserhebung bei auslandischen Kindern und Jugendlichen. Tagungsbericht vom 26.11.1982. Essen (erscheint in Kurze).

Sußmilch, E. 1981. Der Sprachstand auslandischer Grundschuler. Entwicklung eines Wortschatztests und seine empirische Prufung. Unveroffentlichte Diplomarbeit. Dusseldorf: Universitat Dusseldorf.